

1
2 **THE USEFULNESS OF THE SEQUENCE ALIGNMENT METHODS IN**
3 **VALIDATING RULE-BASED ACTIVITY-BASED MODELS**
4

5 **George Sammour, Tom Bellemans, Koen Vanhoof, Davy Janssens, Geert Wets***
6

7
8 Transportation Research Institute (IMOB)
9 Faculty of Applied Economics
10 Hasselt University
11 Wetenschapspark 5 Box 6
12 B-3590 Diepenbeek
13 Belgium
14

15 Fax: +32 11 26 91 99

16 Tel.: +31 11 26 91 58
17

18 Email:

19 {george.sammour,tom.bellemans,koen.vanhoof,davy.janssens,geert.wets}@uhasselt.be
20

21 * Corresponding author
22

23 Number of Words = 5994

24 Number of Tables = 3

25 Number of Figures = 3

26 Words counted: $5994 + 3 * 250 + 3 * 250 = 7494$ words
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43

1 **ABSTRACT**

2
3 The aim of this paper is to achieve a better understanding of rule-based activity-based
4 models, by proposing a new level of validation on the process model level in the
5 ALBATROSS model. To that effect, the work activity process model, which includes six
6 different decision steps, is investigated. Each decision step is evaluated during the prediction
7 of individuals' schedules. The compartment of execution in the process model contains
8 activation dependency. This branches the execution and evaluation of each agent under
9 examination. And yields a sequence of decisions for each agent, where the Sequence
10 Alignment Method (SAM) is employed to evaluate how similar/dissimilar predicted with
11 observed decision sequences are. SAM utterly fits for assessing the analysis of decision
12 sequences on this level. The original CHAID decision trees at each decision step utilized in
13 ALBATROSS are compared with other well known induction methods chosen to appraise
14 the purpose of the analyses. Additionally, the performance of the models is compared at three
15 existing validation levels: the classifier or decision step level using confusion matrix
16 statistics. The work activity trips Origin-Destination (OD) matrix level and time of day work
17 activity start time level, using the correlation coefficient. The results of validation on the
18 proposed process model level show conformance to those already existing, with additional
19 information to help in better understanding the process model's behaviour.
20

21 **INTRODUCTION**

22
23 In the past few decades, many studies have been conducted in order to try to understand the
24 nature of travel demand. Travel demand is derived from the human needs to participate in
25 activities that are distributed in time and space. Models that simulate travel demand using an
26 activity-based approach have been gaining growing attention in recent times due to their
27 strong behavioral foundation and insightful theoretical demand. Recognizing that travel is a
28 demand derived from individuals' needs to perform activities, researchers in travel demand
29 modeling have become increasingly interested in analyzing and predicting individuals'
30 decisions about activity participation. Activity-scheduling models share the objective to
31 predict the sequence of decisions that leads to an observed activity pattern of
32 households/individuals. Activity-based models aim at predicting on a daily basis and for
33 individuals which activities are conducted, by whom, for how long, at what time, the
34 location, and which transport mode is used when traveling is involved (1). The data
35 requirements for activity-based models are in general demanding compared to conventional
36 travel demand models. This is obvious specially that this type of micro-simulation models
37 should be able to predict the travel behaviour in detail including how the activities are
38 selected and scheduled. And so the validation of behavioural models becomes a difficult task.
39 Rule-based activity-based models are no exception as the validation process can be
40 performed on several levels hence, validating the model on an additional or new level may
41 incur extra knowledge to further calibrate and improve its performance. An existing and fully
42 operational rule based activity based model is the ALBATROSS model (1); it is a
43 computational process model, where schedules are predicted using CHAID based induction
44 tree method.

45 The validation of the ALBATROSS model is performed on many different levels, in
46 their original work (1) considered model performance on three levels: (i) the choice facet or

1 the decision tree induction level, by measuring the predictive accuracy of each decision rule
2 in the scheduling process. (ii) At the activity pattern level, sequence alignment methods are
3 used to assess the correspondence between the observed and predicted activity sequences (2).
4 (iii) At the trip matrix level, using correlation coefficients calculated to measure the degree of
5 correspondence between the observed and the predicted Origin-Destination matrices.
6 Decision trees derived from survey data may become large, complex and difficult to
7 interpret. In several experimental and analytical studies using the ALBATORSS model,
8 examples as in (4) and (5) performed validation on three levels, choice facet, activity pattern
9 and trip matrix levels.

10 The objective of this study is to investigate and assess the performance and predictive
11 behaviour of activity-based models on the decision process level of rule based activity based
12 models. The process model level is a core component of the scheduler engine in
13 ALBATROSS, which may reveal extra information on the model. And consequently assess
14 in more understanding the effect of using a specific induction method and in return improve
15 model performance. By further analyzing the process model the sequence alignment method
16 (SAM) was selected to measure how similar predicted to observed decision sequences are.
17 To this end, this work attempts to prove that evaluating activity-based models on this new
18 level expose information helps in additional understanding of the model.

19 The remaining part of this paper is organized as follows, in the next section the
20 ALBATROSS model and the FEATHERS framework used to implement the model for
21 Flanders are described, followed by a discussion of the diary data used for training the
22 model. The analyses and the process model are further discussed explaining the induction
23 methods and elaborating on the usefulness of adapting SAM in process models. Then
24 experiments design and discussion of results are discussed, followed by the conclusion and
25 future works.

27 **THE FEATHERS / ALBATROSS SYSTEM**

28
29 ALBATROSS is a fully operational rule based activity-based model that incorporates
30 household-level decision making (1) (4). In ALBATROSS, rules are used to predict activity-
31 travel choices of individuals and households. The decision rules are formalized from the
32 training of decision trees by using a CHAID decision tree induction method on surveyed
33 activity-travel diary data. In ALBATROSS, to generate a schedule for each person for each
34 day a sequential decision process is assumed, in which the rules are derived from 26 decision
35 trees, and the activity scheduling process model consists of four components or sub models
36 (14). The first component is responsible for generating primary work activities and their start
37 time, duration of each work episode if more than one episode is predicted, and their location,
38 and finally the transport mode for the work trip. The second component is used to generate
39 secondary fixed activities, usually work-related such as bring/get, business or other
40 mandatory activities. In addition it decides which type of activities performed, the number of
41 episodes for each activity, and their start time and duration. The third component is similar to
42 the second component, except it determines the flexible activities part of the schedule. The
43 fourth and last component is in charge of predicting the transport mode of secondary fixed
44 and flexible activities, as the transport mode of the primary work is already decided by the
45 first component. It is important to note that in ALBATROSS the activity travel behavior of
46 the two heads only is captured. A full account of the Albatross model system is given in (1).

1 The analysis performed in this work is performed on the first component dealing with work
2 activity scheduling excluding the transport mode decision step. Figure 1 depicts the work
3 activity decision process model used in ALBATROSS. Each numbered rectangle refers to a
4 decision tree model derived from activity diary data. The index j used in the figure refers to
5 the number of work episodes, if more than one work activity episode is predicted.

6 The first decision step evaluates whether the individual's schedule contains a work
7 activity, if so, the duration of the work activity is predicted next. Followed by the number of
8 work activity episodes, subsequently the ratio between work episodes and the break time
9 duration is decided. And finally the work activity start time is predicted. Decision steps 1 and
10 3 are discrete choice decisions, whereas, decision steps 2, 4, 5 and 6 are continuous choice
11 decisions. It is noteworthy that if decision step 1 infers no work episode for the individual
12 under consideration then decision steps 2-6 will not be executed. Similarly, if decision step 3
13 evaluates to not including a second work episode, then decision steps 4 and 5 will not be
14 evaluated. This implies that there is an activation dependency in the execution of this process
15 model.

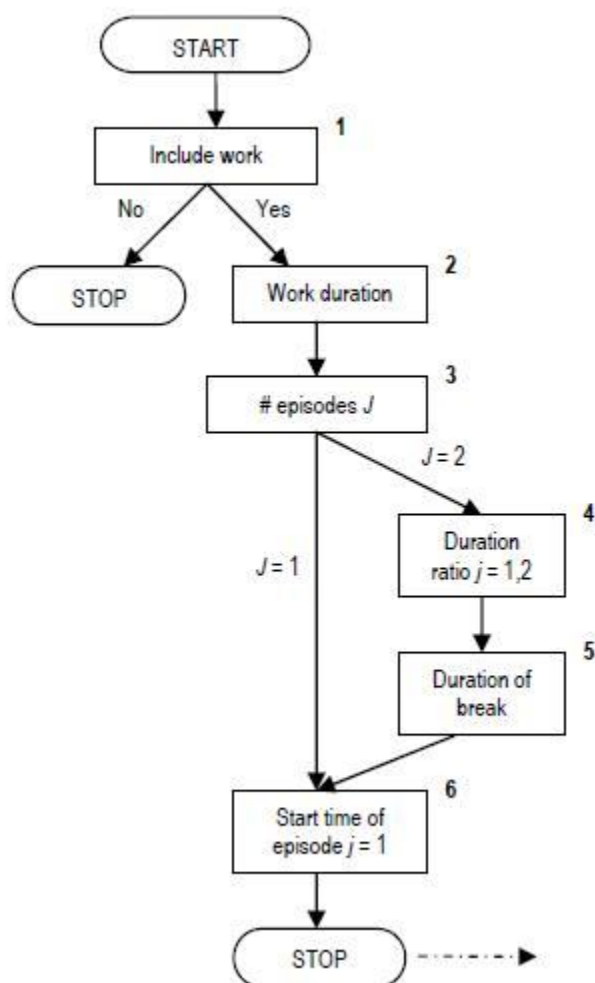
16 The analysis in this work is developed within the FEATHERS (Forecasting Evolutionary
17 Activity-Travel of Households and their Environmental RepercussionS) framework (6). The
18 FEATHERS framework is developed to facilitate the development of modular activity-based
19 models for transportation demand in Flanders (Belgium). The scheduling engine that is
20 currently implemented in the FEATHERS framework is based on the scheduling model that
21 is present in the ALBATROSS system (1). The framework is fully operational at the level of
22 Flanders. The scheduling is based on CHAID decision trees (8), trained based on the
23 Onderzoek VerplaatsingsGedrag Vlaanderen (OVG) travel survey data. The modular design
24 of FEATHERS allows for ease of adaptation of classification methods other than CHAID
25 decision trees, such as Bayesian networks (9), simple classifiers (5), and association rules
26 (10). Taking the above in account, the analysis in this work was conducted based on the
27 ALBATROSS model that was implemented in the FEATHERS framework. However, for
28 research purposes the FEATHERS framework is extended to conduct experiments using
29 alternate induction methods, such as decision trees, logistic regression and OneR (11) (work
30 is still going on to add more methods). This additional functionality allows one to train
31 models outside FEATHERS, using data mining packages that can export Predictive Model
32 Markup Language (PMML) (12). PMML is an XML based language to annotate data mining
33 model parameters in textual form with meta-data for re-use. And thus, using this
34 functionality, the CHAID induction method was replaced by alternatives such as C45 (13),
35 Logistic regression and OneR (12) then integrated within the scheduling model.

37 FLEMISH ACTIVITY TRAVEL DIARY DATA FOR MODEL TRAINING

38
39 The data sets used for training the models in the work activity process model and all the 26
40 decision trees originates from the OVG survey. The survey is a trip-based survey method. The
41 travel survey was conducted based on a random sample from the national register. These persons
42 involved in a survey that was perform primarily through face-to-face interviews. Table 1 shows
43 the situational and socio-demographic variables that are used as prediction variables in
44 FEATHERS/ALBATROSS.

45 The variables that relate to the household level attributes are urban density, household
46 composition, the presence of youngest children in the household, socio-economic class, and
47 car ownership. The gender, driver license, work status and work status of the person's partner

1 are variables related to the individual attributes. In addition, variables such as, the number of
 2 employees with daily-good and non-daily good, number of households within a specific
 3 distance from home location of a household, the distances (in decameters) of the nearest
 4 daily and non-daily good sector and the nearest distance of employees within a ranges are
 5 related to the measures of accessibility given the home location of the household.



37 **FIGURE 1 Work activity process model in ALBATROSS, adapted from (7).**

39 Finally, variables labeled with a (*) are captured and kept from previous decisions and
 40 included in the next decision step, only during the decision process. Continuous variables
 41 such as duration, duration Ratio, break time duration and start time of work activity episodes
 42 are discretised by using Equal Frequency Interval (EFI) method. A 70-30% training-test split
 43 was made on the data. As mentioned above, the datasets for decision steps 1 and 3 are
 44 discrete choice models; with minority class is 28% and 13% respectively.

1 **TABLE 1 Work activity pattern datasets description**

Name	Description	categories
Urb	Urban density	0: highest density, 4: lowest density
Comp	Household composition	0: single without children, 1: single with children, 2: single with parents, 3: partner without children, 4: partner with children
Child	Presence of the youngest children	0: no children, 1: < 6, 2: 6-12, 3: >12 years
Day	Day of the week	0: Monday to 6: Sunday
pAge	Age category	0: <35, 1: 35<55, 2: 55-<65, 3: 65-<75, 4:>75 years
SEC	Household income (in €)	0: <16,250, 1: 16,251 – 23,750, 2: 23,751 – 38,750, 3 >3: 38,750
Ncar	Number of cars in household	0: no cars, 1: 1 car, 2: 2 or more cars
Gend	Gender	0: female, 1: male
Driver	Driving license of person	0: is not a driver, 1: is driver
wstat	Work status of person	0: no work, 1: part time, 2: full time
Pwstat	Work status of person's partner	0: no work, 1: part time, 2: full time
Xdag	Number employees daily-good sector within 3.1 km from home	0: <0,115], 1: <115,253], 2: <253,307], 3: <307,507], 4: <507,675], 5: >675
Xn-dag	Number employees non-daily-good sector within 4.4 km from home	0: <0,395], 1: <395,635], 2: <635,762], 3: <762,938], 4: <938,2525], 5: >2525
Xarb	Number employees within 4.4 km from home	0: <0,8785], 1: <8785,12995], 2: <12995,16120], 3: <16120,20199], 4: <20199,70314], 5: >70314
Xpop	Number households within 3.1 km from home	0: <0,5050], 1: <5050,8845], 2: <8845,13217], 3: <13217,16833], 4: <16833,22884], 5: >22884
Ddag	Distance (dm) to nearest 160 employees daily-good sector	0: <0,71], 1: <71,127], 2: <127,165], 3: <165,202], 4: <202,346], 5: >346
Dn-dag	Distance (dm) to nearest 260 employees non-daily-good sector	0: <0,92], 1: <92,145], 2: <145,176], 3: <176,258], 4: <258,334], 5: >334
Darb	Distance (dm) to nearest 4500 employees total	0: <0,92], 1: <92,128], 2: <128,201], 3: <201,274], 4: <274,360], 5: >360
Dpop	Distance (dm) to nearest 5200 households	0: <0,0], 1: <0,105], 2: <105,126], 3: <126,163], 4: <163,278], 5: >278
Dur*	Total duration (min.) of work activity	0: <0,395], 1: <395,495], 2: <495,526], 3: <526,565], 4: >565
Nep*	Number of work episodes	0: one, 1: two
Ratio*	Ratio (%) between first and second work episodes.	0: <0,40], 1: <40,48], 2: <48,52], 3: <52,60], 4: >60
Inter*	Duration (min.) of break time between first and second work episodes	0: <0,25], 1: <25,47], 2: <47,60], 3: <60,95], 4: >95

2 * Included only if known in stage of the decision process.

3
4 **ANALYSIS**5
6 To be able to analyse the behaviour of the work activity process model only decision steps 1
7 and 3 are replaced by alternative classification methods. Because at these decision steps the
8 execution pattern of the process model is affected. While the continuous decision steps (2, 4,

1 5 and 6) are kept unchanged using the original CHAID based tree induction. The analysis
2 was performed using four different induction methods that are appropriate for assessing the
3 proposed validation level. The first method is the original CHAID tree method. The second
4 technique is the C45 decision tree method for two reasons, (a) C45 is a benchmarking
5 method in the data-mining community, (b) in a case study, Wets et al (16) found
6 approximately equal performance of CHAID and C45 decision tree algorithms in terms of
7 goodness of fit. The third technique is the Logistic Regression classification method, which
8 will be referred to as Logit throughout this paper. The Logit method was selected because it
9 generally outperforms decision tree methods in terms of classification accuracy, especially
10 for small size data sets, as shown by (17). Moreover, Logit can produce probability estimates.
11 The fourth and last method is OneR induction, which is a very simple classifier that provides
12 a rule based on the value of a single attribute. And given the unbalanced nature of the
13 discrete class data sets it is expected that this method will be biased towards the majority
14 class.

15 In the next subsections the induction methods used in the analyses are described,
16 followed by an elucidation of the SAM similarity measure. And in the next section the
17 proposed validation method on the process model level is discussed in details.
18

19 **Decision Tree induction methods general concepts**

20

21 Decision trees are techniques which are used to make decisions from a set of training cases.
22 To use a decision tree for prediction, a rule is specified that assigns a class of the condition
23 attribute to each case classified by the tree. ALBATROSS uses a probabilistic action-
24 assignment rule, for both discrete and continuous choice induction, instead of a deterministic
25 assignment rule, because this results in a better prediction of the aggregate distributions. And
26 so, each rule is assigned a probability distribution that is derived from the frequency
27 distribution over the classes of the condition attribute in the training set for each leaf. An
28 important issue in decision tree learning is over-fitting. The concept of over-fitting occurs
29 when the induction algorithm generates a decision tree that perfectly fits the data in the
30 training data set but lacks the capability of generalization of instances not present in the
31 training set. To avoid over-fitting the minimum number of cases at leaf nodes was set to 30
32 for both CHAID and C45 decision tree models (18).
33

34 **The CHAID decision tree**

35

36 The CHAID was introduced by (8), it originated from the automatic interaction detection
37 (AID) method. The CHAID based induction tree method is able to generate trees with more
38 than two branches attached to the same node at any level of the tree and mainly suited for the
39 analysis of large data sets. It is based on the chi-squared (χ^2) statistic to identify the best split
40 of the data set on condition variables into homogenous partitions with respect to the class
41 variable. In addition the CHAID based tree induction method allows for specifying a
42 threshold (α) for splitting based on the significance level and the minimum number of cases
43 at leaf nodes. The tree building algorithm is performed by recursively iterating through the
44 condition variables to test for each variable the pair of categories whether there is no
45 statistically significant difference within the pair with respect to the class variable. The split
46 with the highest significance value across condition variables is selected. This procedure is

1 repeated until no significant splits are found or the maximum number of cases at leaf nodes is
2 reached.

4 **C45 decision tree**

6 There are two stages for building a classification decision tree in the C4.5 algorithm (11).
7 The first stage involves generating the decision tree based on the training data set, where the
8 second stage has to do with pruning the decision tree based on a validation or test data set
9 that is left out from the training set. The algorithm works as follows. Assume we have a data
10 set S of training cases or samples, where each case consists of n condition or explanatory
11 variables x_1, x_2, \dots, x_n and a class or response variable C_i , for $i = \{1, 2, \dots, p\}$ classes. C4.5 first
12 grows an initial tree using the divide-and-conquer technique by splitting the training set into
13 homogeneous subsets S_1, S_2, \dots, S_p , until the leaf nodes contain only cases from a single
14 class. An important issue in learning classification trees is over-fitting on the data. Therefore
15 to avoid over-fitting C4.5 adopts a pruning strategy, where the decision tree is simplified by
16 removing one or more sub-trees and replacing them with leaves. For a detailed description,
17 the interested reader is referred to (11).

19 **Logistic regression**

21 Logistic regression (18), sometimes referred to as Logit, is an alternative regression
22 technique naturally suited to categorical data. Logit fits an S-shaped curve to the data. Let
23 X, Y be a dataset with a binary response or class variable, where X is a vector of k independent
24 variables (x_1, x_2, \dots, x_k) for each case x_i in X the response or dependent variable is either $y_i=1$
25 or $y_i=0$ then, the logistic model predicts the Logit of Y from X . The Logit is the natural
26 logarithm (\ln) of odds of Y , and odds are ratios of probabilities π of Y happening (i.e., a work
27 activity exists in an individual's schedule at a specific day) to probabilities $(1 - \pi)$ of Y not
28 happening (i.e., a work activity does not exist in an individual's schedule). The simple
29 logistic model has the following form:

$$30 \quad \text{logit}(Y) = \ln\left(\frac{\pi}{1 - \pi}\right) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \quad (1)$$

31 Where \ln is the natural logarithm, π is the probability of the class variable $Y=1$, α is the Y
32 intercept, and $\beta_1, \beta_2, \dots, \beta_k$ are the regression coefficients. The probability (π) that the class
33 variable $Y=1$ is computed by:

$$35 \quad \pi(y = 1) = \frac{\exp(\alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)}{1 - \exp(\alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)} \quad (2)$$

37 The α and $\beta_1, \beta_2, \dots, \beta_k$ are typically estimated by the maximum likelihood method.

39 **One R**

41 One R is a very simple classifier that provides a rule based on the value of a single attribute.
42 According to (12) the algorithm may compete with state-of-the-art techniques used in the
43 field (12). Similar to other algorithms, One R takes as input a set of several attributes and a

1 class variable. Its goal is to infer a rule that predicts the class given the values of the
2 attributes. The One R algorithm chooses the most informative single attribute and bases the
3 rule exclusively on this attribute. Full details can be found in (12). The algorithm assumes
4 that the attributes are discrete. If not, they must be discretised.

6 **THE USEFULNESS OF SAM FOR THE WORK ACTIVITY PROCESS MODEL**

8 **The Sequence Alignment Methods (SAM)**

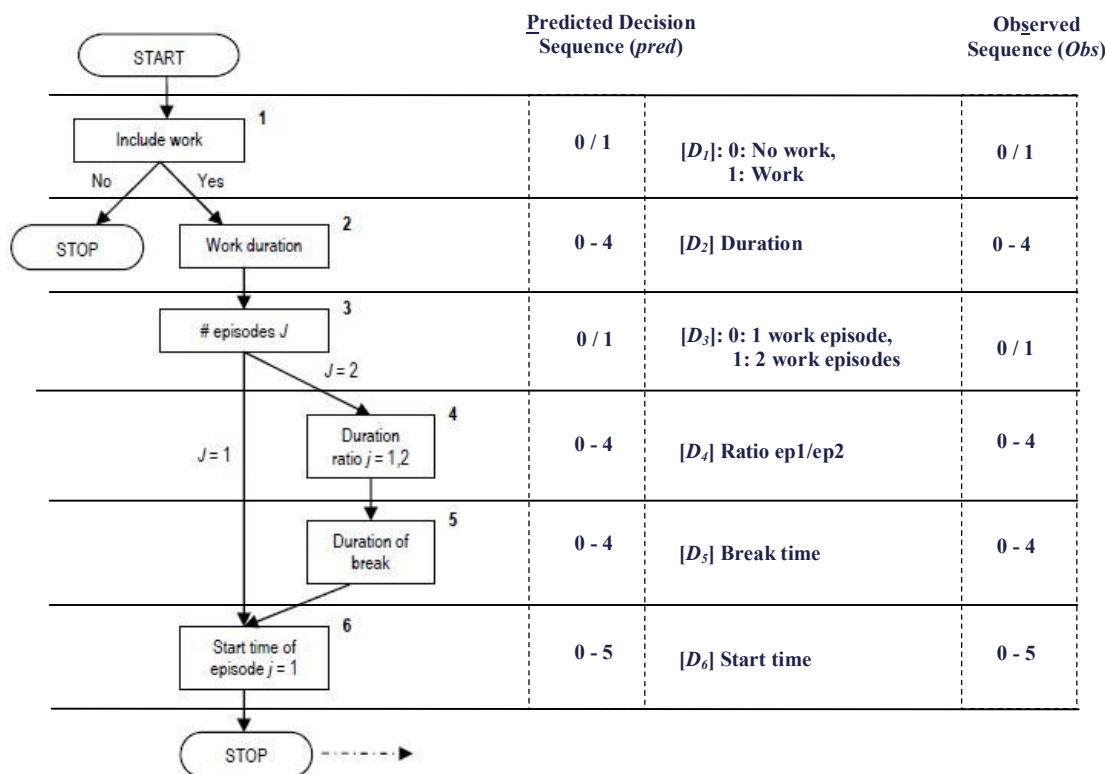
10 The work related to sequential analysis of activity patterns in activity-based models reached a
11 new milestone, when the Sequence Alignment Method (SAM) was introduced in
12 transportation research by Wilson (20). The interesting characteristic of the SAM is that it
13 makes use of biological distance rather than geometric (Euclidean) distance as the basic
14 concept of comparison (21). Mainly in Activity-based models, the SAM methods are used to
15 measure the goodness of fit, in terms of how similar/dissimilar the observed and the
16 predicted activity sequences are. This is done by calculating the effort required to make the
17 two sequences identical using insertion, deletion, and substitution operators. Insertion and
18 deletion operations require the same cost of one unit, whereas substitution requires twice that
19 cost. The lower the SAM measure, the more similar the two sequences are. In the context of
20 this work, the SAM measure will be used on the process model level rather than the activity
21 pattern level. The approach in which the SAM is adopted on the process level and the
22 rationale behind choosing SAM is explained in the next section.

24 **The adaptation of SAM on the decision process model level**

26 The validation of the ALBATROSS model, as mentioned in the introduction, is performed on
27 mainly three levels, the choice facet or the decision tree induction level, the activity pattern
28 level, and the trip (O-D) matrix level. These levels provide goodness-of-fit measures either
29 on individual classifiers, or on the system outputs. However, they do not provide information
30 on the activation dependency and its effect on the model's performance. Therefore, to be able
31 to assess and analyze the behaviour of the decision process model in ALBATROSS, a
32 validation method on the process model level is required. And a measure was needed to
33 appraise the quality of prediction at each decision step. Considering the characteristics of
34 decision outcomes at each decision step in the work activity process model, as shown in
35 Figure 2, the process actually output a *sequence* of decision outcomes or as will be called in
36 the remainder of the paper the decision sequence. Thus, the SAM measure is the best fit for
37 the purpose of assessing the validity on the process model level. The generation of the
38 decision sequences involves the following definitions and assumptions, for each individual:

- 39 - A predicted decision outcome sequence $[D1 D2 D3 D4 D5 D6]_{Pred}$ is generated, and
40 similarly,
- 41 - An observed sequence $[D1 D2 D3 D4 D5 D6]_{Obs}$ that is extracted from diary data is
42 generated accordingly.
- 43 - The length of the predicted and observed sequences can be 1, when no work activity
44 inclusion, 3 when only one work episode is conducted or 6 when two work activity
45 episodes are captured.

1



2

3 **FIGURE 2 Work activity process model decision outcomes in ALBATROSS.**

4

5 Another point of concern, with regard to the proposed level of analysis is the
 6 approach in which the SAM measure is calculated. Will the SAM measure be calculated on a
 7 *one-to-one* (on the single decision step level)? *All-to-all*, taking the whole decision sequence
 8 after the process model finishes execution? Or in a *stepwise* manner, which entails
 9 calculating the SAM after each decision step taking in account the previous decision
 10 outcome, as the process model is executing.

11

12 Using the *one-to-one* approach, the SAM distance will be measured for each decision
 13 step separately. This will serve as an accuracy measure for the individual decision step or the
 14 classifier level itself. Moreover, in the *all-to-all* approach, only one SAM distance is
 15 measured, which indicates how similar the two decision sequences are. Nevertheless, using
 16 this approach will not capture the activation dependency behaviour. And finally, using the
 17 *stepwise* approach, the SAM is evaluated after each decision step keeping the previous
 18 decision. And this entails that at each decision step the SAM distance is measured for the
 19 observed and predicted decision sequences preserving previous decision symbols as the
 20 execution of the process model continues.

21

22 **DESIGN OF EXPERIMENTS AND RESULTS**

23

24 The aim of this study is to validate and assess the performance of activity based models on
 25 the process model level, and further validate that the proposed method on three existing
 validation levels, the classifiers' level, the work activity Origin-Destination (OD) matrices

1 level (spatial resolution), and the work activity start time distribution throughout the day
 2 (temporal resolution). This will allow for assessing the performance of the work activity
 3 process model. The C45 approach was trained using WEKA's J48/C4.5 implementation. The
 4 OneR approach was also trained using WEKA. The Logit models were trained using the
 5 Rattle package for R (22). The models were exported to PMML and a *decisionMaker* class is
 6 implemented in the FEATHERS framework to deploy PMML decision trees as well as Logit
 7 models. The experiments were setup by running FEATHERS for the simulation of cases and
 8 generating schedules for both the training and test sets in four different settings, where in
 9 each setting a different classifier for decision steps 1 and 3 is used for prediction of work
 10 activities in the process model for each day.

12 **Work activity process model level accuracy analysis**

14 The analyses on the process level were conducted by capturing the decision output at each
 15 decision step and calculate the stepwise SAM distance between predicted and observed
 16 decision sequences. This implies that the decision sequence grows in length (depending on
 17 the activation dependency) as the execution of decision steps continues. So the amount of
 18 increase in the SAM distance within the same model approach points out the effect of a
 19 decision step on the previous decision step. The average length of the observed sequence is
 20 1.9 (1.4) symbols with standard deviation between brackets, whereas for CHAID, C45, Logit
 21 and OneR the average lengths are 1.8 (1.5), 1.9 (1.5), 2.04 (1.6), and 1 (0) respectively. The
 22 average length of the OneR approach is 1 with a variance of 0 because the model always
 23 predicts no work activity and hence the decision sequence contains only one symbol. It is
 24 observed that the CHAID and C45 approaches predict similar decision sequence lengths. On
 25 the other hand, the Logit approach predicts longer decision sequences, note that is due to the
 26 activation dependency execution of decisions, and in order to measure the similarity between
 27 predicted and observed decision sequences requires more effort in terms of deletion.

28 Figure 3 depicts the stepwise SAM distance for the training and test sets, at each
 29 decision step in the work activity process model represented in a line chart. The chart
 30 illustrates that the CHAID and C45 approaches reported similar performance and behaviour.
 31 Despite the fact that at decision step 1 the Logit approach performed best, the CHAID and
 32 C45 reported a close decision sequence incremental SAM distance starting from decision
 33 step 2, which means that when evaluating the two steps all together, decision trees
 34 outperform the Logit approach.

36 **Classifier level**

37 **Discrete choice models**

39 The evaluation criteria of the discrete choice models are presented using two accuracy
 40 measures, the confusion matrix (also called contingency table) accuracy measure, since both
 41 discrete choice classifiers are binary. And the Brier score (23) because of the probabilistic
 42 action assignment rule used in scoring the models.

43 The confusion matrix records correctly and incorrectly recognized examples for each class.
 44 The following accuracy statistics can be derived from the confusion matrix:

$$45 \text{ Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (3)$$

1 $Sensitivity = \frac{TP}{TP + FN}$ (4)

2
3 $Specificity = \frac{TN}{TN + FP}$ (5)

4
5 $F - Measure = \frac{2 \times Sensitivity \times Precision}{Sensitivity + Precision}$ (6)

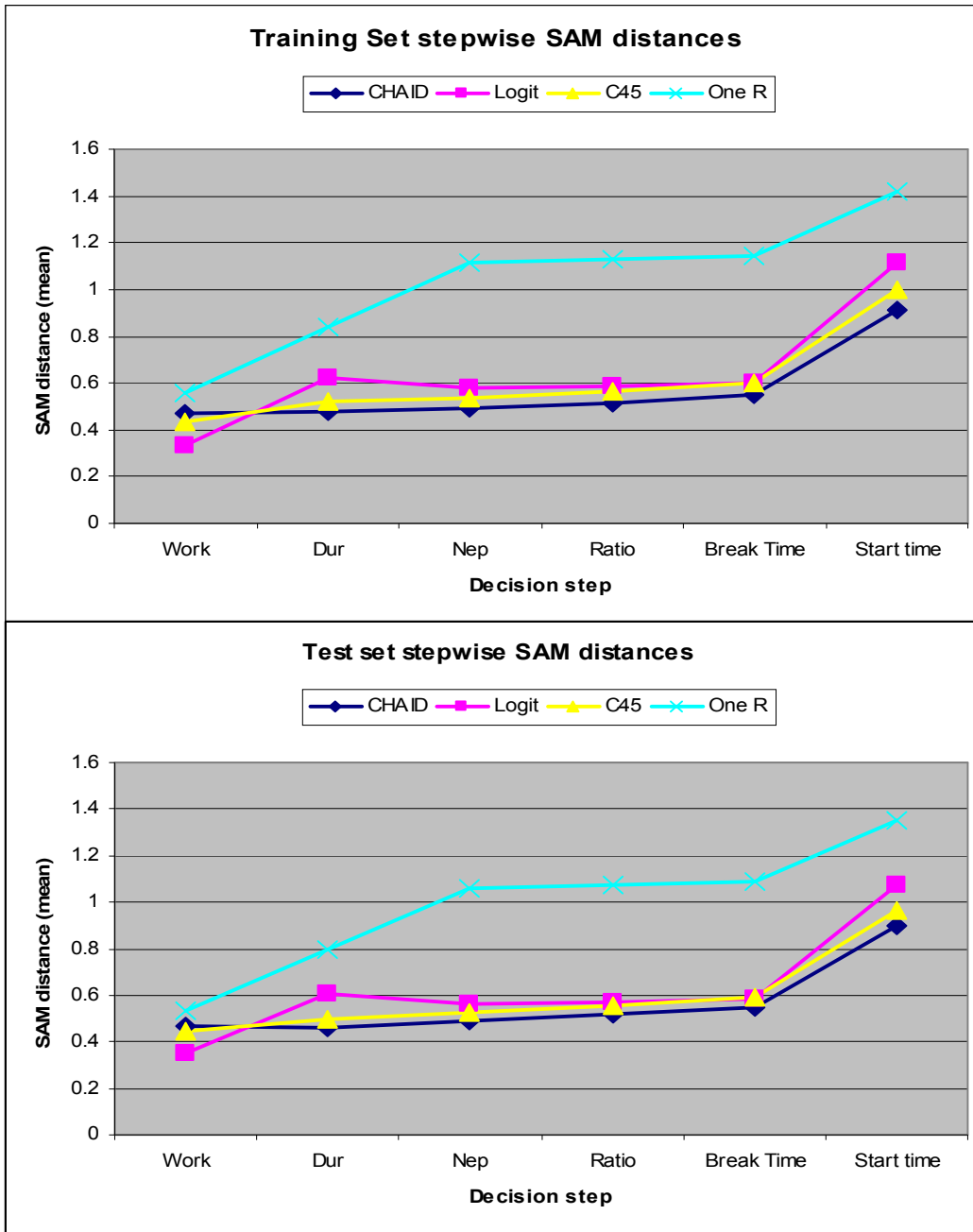


FIGURE 3 Stepwise SAM distances for the work activity process model.

1 Where, TP: number of true positive values, FP: the number of false positive values,
 2 TN: number true negative values and FN: false negative values. The precision in the F-
 3 Measure can be computed as: $precision = TP/(TP+FP)$. Accuracy is not a preferred
 4 performance measure for imbalanced datasets (24). When working with a high imbalance, a
 5 classifier classifying everything as a majority class sample will result in a high predictive
 6 accuracy. Sensitivity approximates the probability of the positive class being correctly
 7 classified, and specificity estimates the probability of correctly predicting the negative class.
 8 The F-measure focuses more on the dropout class by consideration of sensitivity and
 9 precision as it is the weighted average of the precision and recall. An F-measure value
 10 reaches its best value at 1 and its worst value at 0.

11 The Brier score (BS) is a metric related to the mean-squared-error often used in statistical
 12 fitting as a measure of model goodness. It is a descriptive measure often used in the literature
 13 on prediction accuracy. The Brier score is calculated as follows:

$$15 \quad BS = \frac{1}{N} \sum_{i=1}^N (p_i - o_i)^2 \quad (7)$$

16
 17 Where p_i is the predicted probability and o_i is the observed value of the instance i (0 if
 18 negative and 1 if positive). The BS measures the average squared deviation between
 19 predicted probabilities for a set of events and their outcomes. So a lower score represents a
 20 higher accuracy. Table 2 provides the results of the analysis to assess model performance.
 21 The results suggest that for decision step 1, the Logit model outperforms all other methods
 22 specially in predicting the positive class value (yWo). As expected, CHAID and C45 show
 23 similar performance with a slight increase in performance in favor of C45. The predictive
 24 performance (sensitivity) for the (yWo) class variable, which is the minority class, is notably
 25 higher in the Logit approach and this can be explained by the fact that Logit outperforms
 26 decision tree approaches for small size datasets (18). The OneR approach prediction outcome
 27 was always no work since the distribution of the class variable in this dataset is skewed (72%
 28 no work), with a Brier score equal to the percentage of the minority class in the dataset.
 29 Results also suggest that the drop in the accuracy in the test set was not significant, while
 30 there was a slight increase in accuracy for the CHAID approach.

31 Considering the performance of decision step 3, again CHAID and C45 confirmed
 32 similar performance but outperform the Logit and OneR approaches, the reason for the
 33 weaker performance of the Logit approach is that the data set at decision step 3 is highly
 34 skewed 87% and this leads to underestimating the rare class calculated by Equation 2 as
 35 reported by (24). Finally the OneR model always predicts the majority and so the predictive
 36 power of the minority class is zero. The NA in the OneR approach indicates that the measure
 37 cannot be computed since the TP and FP values used to calculate the precision for this
 38 approach are zero.

40 Continuous choice models

41
 42 The continuous choice models where trained using only the CHIAD tree induction method
 43 used originally in ALBATROSS were kept the same for the analyses performed using
 44 alternative discrete choice models. The performance of continuous choice models was
 45 assessed by means of the Relative Absolute Error (RAE) which gives an indication of how

1 good a predicted value is relative to the observed value. The reason for selecting this measure
 2 is that it can be reported as a percent error measure for numeric or continuous predictions.
 3 The RAE is calculated by dividing the sum of the absolute difference between the predicted
 4 and observed values by the observed cases. Results showed fairly good results with 21%,
 5 22.4% and 9% for decision steps 2,4 and 6 respectively for training sets, and 20%, 20.4% and
 6 10% for test sets, while for decision step 5 the RAE reported 64% for training and 61% for
 7 test set.

8
 9 **TABLE 2 Accuracy statistics for discrete choice models (classifier level)**

Work	Training set			
Model	Brier Score	Sensitivity	Specificity	F-Measure
CHAID	0.11766	0.54065	0.841026	0.554455
Logit	0.113781	0.813008	0.839448	0.73026
C45	0.114957	0.59248	0.84497	0.594898
OneR	0.279625	0	1	NA
Work	Test set			
Model	Brier Score	Sensitivity	Specificity	F-Measure
CHAID	0.112366	0.554371	0.851653	0.563991
Logit	0.115959	0.791045	0.83628	0.704653
C45	0.115108	0.556503	0.825519	0.545455
OneR	0.264972	0	1	NA
Nep	Training set			
Model	Brier Score	Sensitivity	Specificity	F-Measure
CHAID	0.106202	0.1875	0.890315	0.195122
Logit	0.0979974	0.125	0.985998	0.205128
C45	0.108202	0.242188	0.866978	0.227106
OneR	0.129949	0	1	NA
Nep	Test set			
Model	Brier Score	Sensitivity	Specificity	F-Measure
CHAID	0.136115	0.155844	0.885496	0.179104
Logit	0.144773	0.0649351	0.959288	0.102041
C45	0.134237	0.194805	0.903308	0.230769
OneR	0.16383	0	1	NA

10
 11 **Work activity trip matrix and trips start time level accuracy analysis (spatial and**
 12 **temporal resolutions)**

13
 14 At the work activity trip matrix level (spatial resolution), the observed and predicted OD
 15 matrices, for training and test sets, were compared. An activity OD matrix contains the
 16 frequency of work activity trips for each combination of origins (rows) and destinations
 17 (columns). The frequency of trips at each zone in Flanders was aggregated forming a one
 18 dimensional array with work activity trip counts at each zone. The correlation is calculated
 19 between observed and predicted matrix entries $\rho(\text{observed}, \text{predicted})$.

20 The work activity start time level (temporal resolution) was also analysed by
 21 calculating the correlation between the observed and predicted work activity start times for
 22 each hour of the day, the reason this analysis was conducted to further investigate the larger
 23 increase in the SAM distance on the process model level at decision step 6 (*Start time*) as can

1 be observed in Figure 2. The indication of NA in Table 3 for the OneR approach indicates
 2 that the correlation is not available since no work activities were predicted using this
 3 approach. The results in Table 3 indicate that the correlation coefficients are similar with the
 4 Logit approach having a slightly lower correlation coefficient than the CHAID and C45
 5 approaches.

7 **TABLE 3 Work activity trip matrix and time of day correlation coefficients**

Dataset	Work activity trip matrix level				Work activity start time per hour of the day			
	CHAID	Logit	C45	OneR	CHAID	Logit	C45	OneR
Training	0.832	0.802	0.835	NA	0.896	0.873	0.899	NA
Test	0.816	0.799	0.82	NA	0.827	0.771	0.803	NA

9 CONCLUSION AND FUTURE WORK

11 From a data mining perspective, rule based activity based models are validated on mainly
 12 three major levels, namely, the classifier accuracy level, where single rules are evaluated and
 13 analysed, the generated activity pattern level using the sequence alignment (SAM) distance
 14 measure by calculating how similar the observed and the predicted activity sequences are.
 15 And the trip matrix level by assessing the correlation coefficient to measure the degree of
 16 correspondence between the observed and the predicted origin-destination matrices. The
 17 work reported in this study proposed a methodology to validate rule-based activity-based
 18 models on the process model level. The proposed analyses suggested that conducting an
 19 investigation on the process model level, provides additional information on how the model
 20 performs when using a specific classifier at a specific decision step. The results obtained
 21 from the analyses, conform to other levels of validation. Plus extra information indicating
 22 that, despite the fact that a classifier's predictive performance is compelling, yet the
 23 activation dependency of the process model affects the overall model performance and
 24 accuracy. Additionally, the results showed that the branching of decision steps at 1 and 3 is a
 25 critical issue for the outcome of the model. And perhaps changing the order of such decisions
 26 might lead to a better model.

27 Future work will be directed towards approaches related to changing the order of decision
 28 steps in the process model. And training the models with and without the inclusion of
 29 additional features in the training data sets in subsequent models and investigate the
 30 performance and behaviour of the model for each setting.

32 REFERENCES

- 34 1. Arentze, T.A., and Timmermans, H.J.P. ALBATROSS: A Learning-based
 35 Transportation Oriented Simulation System. EIRASS, Eindhoven University of
 36 Technology, The Netherlands, 2000.
- 37 2. Joh, C., H. Arentze, T.A. and Timmermans, H.J.P. Pattern Recognition in Complex
 38 Activity-Travel Patterns: A Comparison of Euclidean Distance, Signal Processing
 39 Theoretical, and Multidimensional Sequence Alignment Methods. Presented at the
 40 80th Annual Meeting of the Transportation Research Board, Washington, D.C., USA,
 41 2001.

- 1 3. Arentze, T.A., and Timmermans, H.J.P. Measuring Impacts of Condition Variables in
2 Rule-Based Models of Space-Time Choice Behavior: Method and Empirical
3 Illustration, *Geographical Analysis*, 2003, 35, 24-45.
- 4 4. Janssens, D., G. Wets, T. Brijs, K. Vanhoof, T. A. Arentze, and H. J. P. Timmermans.
5 Integrating Bayesian Networks and Decision Trees in a Sequential Rule-Based
6 Transportation Model. *European Journal of Operational Research*, Vol. 175, No. 1,
7 2006, pp. 16–34.
- 8 5. Moons, E. Modelling Activity-Diary Data: Complexity or Parsimony? PhD
9 dissertation. Limburg University, Diepenbeek, Belgium, 2005.
- 10 6. Bellemans T., Janssens D., Wets G., Arentze, T., and Timmermans H..
11 Implementation Framework and Development Trajectory of the FEATHERS
12 Activity-Based Simulation Platform. Proceedings of the Annual Meeting of the
13 Transportation Research Board, 2010
- 14 7. Arentze, T.A., and Timmermans, H.J.P. A Learning-Based Transportation Oriented
15 Simulation System. *Transportation Research Part B*, 38, 2004, 613-633.
- 16 8. Kass, G.V. (1980) An exploratory technique for investigating large quantities of
17 categorical data. *Applied Statistics*, 29, 119{127.
- 18 9. Janssens, D., G. Wets, T. Brijs, K. Vanhoof, T. A. Arentze, and H. Timmermans.
19 Improving Performance of a Multiagent Rule-Based Model for Activity Pattern
20 Decisions with Bayesian Networks. In *Transportation Research Record: Journal of*
21 *the Transportation Research Board*, No. 1894, Transportation Research Board of the
22 National Academies, Washington, D.C., 2004, pp. 75–83.
- 23 10. Keuleers, B., G. Wets, T. Arentze, and H. Timmermans. Association Rules in
24 Identification of Spatial-Temporal Patterns in Multiday Activity Diary Data. In
25 *Transportation Research Record: Journal of the Transportation Research Board*, No.
26 1752, TRB, National Research Council, Washington, D.C., 2001, pp. 32–37.
- 27 11. Holte, R. C.. Very simple classification rules perform well on most commonly used
28 datasets. *Machine Learning*, 1993, 11(1):63–90.
- 29 12. Guazelli, Alex, Wen-Ching Lin and Tridivesh Jena. *PMML in Action: Unleashing the*
30 *Power of Open Standards for Data Mining and Predictive Analytics*. CreateSpace.
31 ISBN 978-1452858265, 2010, pp 5.
- 32 13. Quinlan, J.R. *Induction of Decision Trees*, *Machine Learning*, vol. 11, no. 1, 1986,
33 pp. 81–106.
- 34 14. Anggraini R., Arentz T., and Timmermans H. Modeling car allocation decisions in
35 automobile deficient households, in: *Proceedings of the European transport*
36 *conference*, Noordwijkerhout, The Netherlands, 2007.
- 37 15. Wets G, Vanhoof K, Arentze T A, Timmermans H J P, Identifying decision structures
38 underlying activity patterns: an exploration of data mining algorithms" *Transportation*
39 *Research Record* number 1718, 2000, 1 – 9.
- 40 16. Lim, T.S., Loh, W.Y. and Shih, Y.S. A Comparison of Prediction Accuracy,
41 Complexity, and Training Time for Thirty-three Old and New Classification
42 Algorithms" *Machine Learning*, 2000, 40, 203-228.
- 43 17. D. R. Cox. The regression analysis of binary sequences. *Journal of the Royal*
44 *Statistical Society, Series B (Methodological)*, 20(2): 1958, pp 215–242.

- 1 18. Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian
2 H. Witten; The WEKA Data Mining Software: An Update; SIGKDD Explorations,
3 Volume 11, Issue 1, 2009.
- 4 19. Wilson, C. Activity Pattern Analysis by Means of Sequence-Alignment Methods.
5 Environment and Planning A, Vol. 30, 1998 , pp. 1017–1038.
- 6 20. Joh, C-H, Arentze, T.A., Hofman, F. and Timmermans, H.J.P. Activity-travel pattern
7 similarity: a multidimensional alignment method. Transportation Research B, 2002,
8 36, 385-403.
- 9 21. Williams, G. J. Rattle: A Data Mining GUI for R. The R Journal, 1(2), 2009, 45-55
- 10 22. Brier, G. W. Verification of forecasts expressed in terms of probability. Mon. Wea.
11 Rev., 1950, 78, 1–3.
- 12 23. King, G. and Zeng, L. Logistic regression in rare events data. Political Analysis,
13 2001, 9:137-163.